

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 5月23日

出 願 番 号

Application Number:

特願2001-153345

[ST.10/C]:

[JP2001-153345]

出 願 人

Applicant(s):

株式会社日立製作所

USSN 10/082,303

Mattingly Stanger Malur

703 684-1120

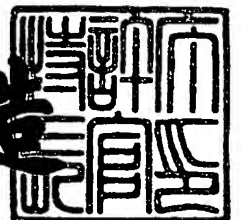
Date ASA-1068

CERTIFIED COPY OF
PRIORITY DOCUMENT

2002年 3月22日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2002-3019586

【書類名】 特許願

【整理番号】 K00020371A

【あて先】 特許庁長官

【国際特許分類】 G06F 3/06

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

 【氏名】 本田 聖志

【発明者】

 【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージ事業部内

 【氏名】 齊木 栄作

【発明者】

 【住所又は居所】 神奈川県小田原市国府津 2 8 8 0 番地 株式会社日立製作所 ストレージ事業部内

 【氏名】 高安 厚志

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100075096

 【弁理士】

 【氏名又は名称】 作田 康夫

【手数料の表示】

 【予納台帳番号】 013088

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

【物件名】 要約書 1
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶装置システム

【特許請求の範囲】

【請求項 1】

上位装置と接続された複数の記憶装置からなる記憶装置システムにおいて、
前記複数の記憶装置に含まれる第 1 の記憶装置が、
前記上位装置における前記記憶システムに対する情報処理についての要求を受信する手段と、

受信された前記要求を、前記複数の記憶装置に含まれる第 2 の記憶装置に転送可能な手段と、

受信された前記要求が、当該第 1 の記憶装置が実行すべき要求である場合に、
受信された前記要求が示す情報処理を実行する手段とを有することを特徴とする記憶装置システム。

【請求項 2】

請求項 1 に記載の記憶装置システムにおいて、

前記情報処理を実行する手段は、当該第 1 の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を実行すべきと判断された場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項 3】

請求項 2 に記載の記憶装置システムにおいて、

前記要求には、前記情報処理を実行すべき記憶装置を示す第 1 の識別情報が含まれ、前記強調制御情報には、当該第 1 の記憶装置を識別する第 2 の識別情報が含まれ、

前記情報処理を実行する手段は、前記第 1 の識別情報と前記第 2 の識別情報が対応する場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項 4】

請求項 1 乃至 3 のいずれかに記載の記憶装置システムにおいて、
前記第 2 の記憶装置が、

前記転送される要求を受信する手段と、

転送された前記要求が、当該第 2 の記憶装置で実行すべき要求である場合に、転送された前記要求が示す情報処理を実行する手段とを有することを特徴とする記憶装置システム。

【請求項 5】

請求項 4 に記載の記憶装置システムにおいて、

前記第 1 の記憶装置では、転送可能な手段が、転送する要求に当該第 1 の記憶装置を示す情報を付加し、

前記第 2 の記憶装置は、付加された前記第 1 の記憶装置を示す情報に基づいて、転送された前記要求の再度の転送を抑止する手段を、さらに有することを特徴とする記憶装置システム。

【請求項 6】

請求項 1 に記載の記憶装置システムにおいて、

前記転送可能な手段は、当該第 1 の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を実行すべきでないと判断された場合に、受信された前記要求を、前記第 2 の記憶装置に転送することを特徴とする記憶装置システム。

【請求項 7】

請求項 1 または 6 のいずれかに記載の記憶装置システムにおいて、

前記転送可能な手段は、当該第 1 の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を前記第 2 の記憶装置が実行すべきと判断された場合に、受信された前記要求を、前記第 2 の記憶装置に転送することを特徴とする記憶装置システム。

【請求項 8】

請求項 7 に記載の記憶装置システムにおいて、

前記要求には、前記情報処理を実行すべき記憶装置を示す第 1 の識別情報が含まれ、前記強調制御情報には、当該第 1 の記憶装置を識別する第 2 の識別情報が含まれ、

前記情報処理を実行する手段は、前記第 1 の識別情報と前記第 2 の識別情報が

対応する場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項 9】

請求項 1 および 6 乃至 8 のいずれかに記載の記憶装置システムにおいて、

前記第 2 の記憶装置が、

前記転送される要求を受信する手段と、

転送された前記要求が、当該第 2 の記憶装置で実行すべき要求である場合に、
転送された前記要求が示す情報処理を実行する手段とを有することを特徴とする
記憶装置システム。

【請求項 10】

請求項 9 に記載の記憶装置システムにおいて、

前記第 1 の記憶装置では、転送可能な手段が、転送する要求に当該第 1 の記憶
装置を示す情報を付加し、

前記第 2 の記憶装置は、付加された前記第 1 の記憶装置を示す情報に基づいて
、転送された前記要求の再度の転送を抑止する手段を、さらに有することを特徴
とする記憶装置システム。

【請求項 11】

請求項 1 乃至 10 のいずれかに記載の記憶装置システムにおいて、

前記要求は、前記複数の記憶装置のいずれかに記憶された情報に対する読み出
し要求および前記複数の記憶装置のいずれかへの情報の書き込み要求のうち少な
くとも一方が含まれることを特徴とする記憶装置システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、データを二重化し保持する記憶装置システムに係る。特に、当該記
憶装置システムを構成する複数の記憶装置間での協調制御に関する。

【0002】

【従来の技術】

“A Case for Redundant Array of Inexpensive Disks (RAID)” (Proceedings
of ACM SIGMOD, 1988)には、複数のディスク装置から構成されるディスクアレ

イ装置における、冗長データの生成、及び格納方法による分類が規定されている。

【0003】

上記の技術によれば、二台のディスク装置に対して二重化したデータを格納することで、冗長構成を取る一台のディスク装置において障害が発生し、当該ディスク装置の格納データの読出し、或いは、当該ディスク装置に対する書込みが不可能になった場合でも、他のディスク装置を用いてデータの読出し、書込みを可能とするディスクアレイ装置は、RAID1と呼んでいる。

【0004】

更に、上記RAID1のディスクアレイ装置において、ディスクアレイ装置内のディスク制御装置から二重化したディスク装置に対するライトデータの転送処理を軽減することで、ディスク制御装置の処理性能を向上させる技術が、特開平10-74129（第一の従来技術）に開示されている。

【0005】

第一の従来技術によれば、データライトの際、ディスク制御装置は、第一のディスク装置にのみライトデータを転送し、該ライトデータを第一のディスク装置が、ディスク制御装置を経由せず、第二のディスク装置に転送することによって、ディスク制御装置から第二のディスク装置に対するライトデータの転送処理を不要とすることが可能となる。

【0006】

更に、二台のディスク装置間を専用の連絡線で相互に結合し、データライトの際、相互に連絡を取りながら同時にライト処理を実施する方法が、特開平7-281959（第二の従来技術）に開示されている。

【0007】

第二の従来技術によれば、二台のディスク装置に対し同時にライトデータの転送を実施することによって、インタフェース上のライトデータの転送量は、二重化しない場合と同等で実現することが可能となる。

【0008】

【発明が解決しようとする課題】

上述の第一の従来技術においては、少なくとも、ディスクアレイ装置を構成する個々のディスク装置を制御し、第一のディスク装置に対して特殊なライトコマンドを発行する手段を持つディスク制御装置を必要とする為、ディスクアレイ装置が高価になるという課題がある。

【 0 0 0 9 】

更に、上記従来技術においては、インタフェース上のライトデータの転送量を低減するものではなく、即ち、ディスク制御装置とディスク装置とを接続するインタフェース負荷を軽減するものではない為、ディスクアレイ装置内のインタフェースの競合による性能低下を改善することができないという課題がある。

【 0 0 1 0 】

また、第二の従来技術においては、第一の従来技術で課題とされた、ディスク制御装置を不要とし、更に、インタフェース上のライトデータの転送量増加を防止するものであるが、その実現において、ディスク装置間で連絡をとりあう為の専用の連絡線が必要となる。

【 0 0 1 1 】

更に、二台のディスク装置で同期して、或いは、主側装置が副側装置の動作の進行を確かめながら、ライトデータの転送処理を実現する為、ディスク装置の利用効率の低下による性能低下、或いは、一台のディスク装置のデータライト処理に対し余分に処理時間が掛かることによる性能低下という課題がある。

【 0 0 1 2 】

また、上述の従来技術においては、第一／第二、或いは、主側／副側の切り替え制御方法について、更に、データリード処理時のディスク装置選択制御方法について、充分考慮されているとは言えない。

【 0 0 1 3 】

本発明の目的は、複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、上位装置からのデータリード／ライト要求を、該記憶装置システムを構成する複数の記憶装置間で協調して処理する制御方式を提供することにある。このことにより、上述の課題を解決する記憶装置システムを提供することである。

【 0 0 1 4 】

【課題を解決するための手段】

上記目的を達成するために、複数の記憶装置から構成されるRAID1を含む記憶装置システムにおいて、以下の構成をとる。

【0015】

本発明では、記憶装置システムに含まれる記憶装置が、上位装置における記憶装置システムに対する情報処理についての要求を受信する手段と、受信された要求を、複数の記憶装置に含まれる第2の記憶装置に転送可能な手段と、受信された要求が、当該第1の記憶装置が実行すべき要求である場合に、受信された要求が示す情報処理を実行する手段とを有するものである。

【0016】

また、記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対応する記憶装置との間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備する構成であってもよい。

【0017】

【発明の実施の形態】

以下、本発明に係る実施形態を、図を用いて説明する。

図1は、二台の記憶装置1-a/bから構成される記憶装置システム4と、複数の上位装置2とが、任意のインタフェース（図中、SAN: Storage Area Network）3で接続される、情報処理システムの一構成例を示す図である。

【0018】

同図において、記憶装置1としてディスク装置を例に構成を示しており、該ディスク装置1-a/bは、夫々前記上位装置2がアクセスするデータを保持する記憶媒体であるディスク部11と、該ディスク部11と前記上位装置2との間で転送されるデータを一時保持するバッファ部12と、前記上位装置2との間のインタフェースプロトコル制御を実行するインタフェース制御部13と、前記ディスク部11に対する記録再生処理を実行するディスク制御部14と、前記バッファ部12に対するアクセスを制御するバッファ制御部15と、上記各部位を統括制御

するメイン制御部 1 0 とから構成されている。

【0 0 1 9】

また、同図において、前記ディスク装置 1 -a/b が二重化されたデータを保持することによって、前記 RAID1 の記憶装置システムを構成するものである。

【0 0 2 0】

尚、同図において、該記憶装置システムを二台のディスク装置から構成する場合を例としているが、本発明の記憶装置システムの構成は、これに限るものではなく、3 台以上の記憶装置が含まれてもよい。

【0 0 2 1】

以下、本発明に係る第一の実施形態を、図 2、3、4、5、6、7、8 を用いて説明する。本実施形態は、各記憶装置が要求を有し（もしくは受信し）、各記憶装置が当該記憶装置でその要求に対応する情報処理を実行すべきか判断する。判断には、協調制御情報 2 8 を用いる。

【0 0 2 2】

本実施形態では、上位装置 2 から記憶装置システム 4 に対するアクセス要求を、該記憶装置システム 4 を構成する複数の記憶装置 1 で共有し、更に、個々の記憶装置において、前記アクセス要求を自身が処理すべきか否かを判別することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現することを可能にする。

【0 0 2 3】

図 2 は、前記記憶装置システム 4 を構成するディスク装置 1 の一構成例を示す図である。

同図において、前記インタフェース制御部 1 3 は、前記インタフェース 3 を介して情報の受信を実施する受信部 2 0 と、同様に情報の送信を実施する送信部 2 1 と、受信したフレーム等の情報に対するエラー検出、或いは、該フレーム情報の少なくとも一部を前記バッファ部 1 2 に格納する際、格納先の制御等を実施する受信フレーム処理部 2 2 と、フレーム等の情報送信の際、該フレーム情報を構成するヘッダ情報等の付加情報の生成を実施する送信フレーム生成部 2 3 とを具

備する。

【0024】

更に、前記受信部20で受信したフレーム等の情報を後段の装置に対し再送するか否かを判別し、再送制御信号を生成する再送判定部24と、上記再送制御信号に基づき、前記送信フレーム生成部23からの情報と、前記受信部20で受信した情報との一方を選択し、前記送信部21に出力する出力選択部25とを具備する。

【0025】

また、前記バッファ部12は、上位装置2から受信したアクセス要求（コマンド）を保持する受信コマンド格納部26と、上位装置2との間で送受信されるデータを保持する送受信データ格納部27とを具備する。

【0026】

また、前記メイン制御部10では、上位装置2から記憶装置システム4に対するアクセス要求について、該アクセス要求を後段の装置に再送するか否か、更には、該アクセス要求を自身が処理すべきか否かを判別する為の情報等、対を成すディスク装置との間での協調制御情報28を保持している。なお、協調制御情報28の詳細に関しては、図16を用いて後述する。

【0027】

更にまた、前記再送判定部24では、上位装置2から記憶装置システム4に対するフレーム等の情報の再送制御として、アクセス要求については、前記協調制御情報28に基づく再送制御を、ライトデータについては、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を実施するものである。

【0028】

以下、上位装置から記憶装置システムに対するアクセス要求について、アクセス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを個々のディスク装置で判別する場合のリード／ライトデータ転送処理を例に動作を説明する。

【0029】

図 3 は、前記ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、上位装置 2 からのリードデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【 0 0 3 0 】

(1) アクセス要求 (コマンド) 受信及び再送処理

上位装置 2 から記憶装置システム 4 に対して発行されたリードデータ要求は、フレーム情報として、先ず、ディスク装置 1-a の前記受信部 20 を介して受信される。ディスク装置 1-a では、前記リードデータ要求を、前記受信フレーム処理部 22 を介して、前記バッファ部 12 の前記受信コマンド格納部 26 に格納する。

【 0 0 3 1 】

また、ディスク装置 1-a の前記再送判定部 24 は、前記協調制御情報 28 に基づき、上位装置 2 から記憶装置システム 4 に対するアクセス要求を後段のディスク装置 1-b に対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部 25 に対する再送制御 (指示) 信号を生成する。結果、前記出力選択部 25 は、上記再送制御 (指示) 信号に基づき、前記上位装置からのリードデータ要求を後段のディスク装置 1-b に対し再送する。

【 0 0 3 2 】

ディスク装置 1-b では、前記ディスク装置 1-a によって再送された前記リードデータ要求を、前記受信フレーム処理部 22 を介して、前記バッファ部 12 の前記受信コマンド格納部 26 に格納する。また、ディスク装置 1-b の前記再送判定部 24 は、前記協調制御情報 28 に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【 0 0 3 3 】

(2) アクセス要求 (コマンド) 解釈及びアクセス要求 (コマンド) 実行

ディスク装置 1-a/b は、前記受信コマンド格納部 26 に格納したアクセス要求を解釈し、該アクセス要求がリードデータ要求であることを検出する。

【 0 0 3 4 】

更に、該リードデータ要求を構成するリード要求領域情報と、前記協調制御情報 2 8 とに基づき、前記該リードデータ要求を自身が処理すべきか否かを判別する。

【 0 0 3 5 】

自身が処理すると判断したディスク装置では、前記ディスク制御部 1 4 を介してディスク部 1 1 から前記バッファ部 1 2 の前記送受信データ格納部 2 7 に対するディスクリード処理を開始し、更に、前記インタフェース制御部 1 3 を介して上位装置 2 に対するリードデータ送信処理を実施する。更に、リードデータ送信後、前記リードデータ要求に対するステータス情報を生成送信し、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【 0 0 3 6 】

一方、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、リードデータ要求に掛かる処理を完了する。

【 0 0 3 7 】

なお、本実施形態では、アクセス要求が、ディスク装置 1-a を介してディスク装置 1-b に転送されるが、ディスク装置それぞれに、上位装置 2 がアクセス要求を送信してもよい。

【 0 0 3 8 】

図 4 は、前記ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、上位装置 2-a/b からのリードデータ要求が発行された場合、特に、上位装置 2-a からのリードデータ要求がディスク装置 1-a で、上位装置 2-b からのリードデータ要求がディスク装置 1-b で処理される場合の、リードデータ要求、及び、リードデータの転送処理等の流れを示す図である。

【 0 0 3 9 】

上述の説明の通り、上位装置 2-a/b からのリードデータ要求は、少なくともディスク装置 1-a のインタフェース制御部 1 3 を介して、ディスク装置 1-a/b のバッファ部 1 2 に格納される。

【 0 0 4 0 】

ディスク装置 1-a/b では、受信したアクセス要求について、自身が処理すべき

可否かを判別し、自身が処理すると判断したリードデータ要求に対して、ディスク部 1 1 からのディスクリード処理、及び、上位装置 2 -a/b に対するリードデータ送信処理、更に、ステータス生成送信処理を実施する。ここで、ディスク装置 1 -a が送信するリードデータは、上位装置 2 -a に対するものであることから、従来通りディスク装置 1 -b のインタフェース制御部 1 3 を介して上位装置 2 -a に転送される。

【0041】

図 5 は、前記ディスク装置 1 -a/b から構成される記憶装置システム 4 に対し、上位装置 2 からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【0042】

(1) アクセス要求（コマンド）受信及び再送処理

対象となるアクセス要求が、リードデータ要求でなくライトデータ要求であること以外は、図 3 を用いて上述した内容と同じである。上位装置 2 から記憶装置システム 4 に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置 1 -a の前記受信部 2 0 を介して受信される。

【0043】

ディスク装置 1 -a では、前記ライトデータ要求を、前記受信フレーム処理部 2 2 を介して、前記バッファ部 1 2 の前記受信コマンド格納部 2 6 に格納する。

【0044】

また、ディスク装置 1 -a の前記再送判定部 2 4 は、前記協調制御情報 2 8 に基づいて、上位装置 2 から記憶装置システム 4 に対するアクセス要求を後段のディスク装置 1 -b に対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部 2 5 に対する再送制御（指示）信号を生成する。結果、前記出力選択部 2 5 は、該再送制御（指示）信号に基づき、前記上位装置からのライトデータ要求を後段のディスク装置 1 -b に対し再送する。

【0045】

ディスク装置 1 -b では、前記ディスク装置 1 -a によって再送された前記ライト

データ要求を、前記受信フレーム処理部 22 を介して、前記バッファ部 12 の前記受信コマンド格納部 26 に格納する。また、ディスク装置 1-b の前記再送判定部 24 は、前記協調制御情報 28 に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【0046】

(2) アクセス要求 (コマンド) 解釈及びアクセス要求 (コマンド) 実行

対象となるアクセス要求が、リードデータ要求でなくライトデータ要求であること以外は、図 3 を用いて上述した内容と同じである。ディスク装置 1-a/b は、前記受信コマンド格納部 26 に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出する。更に、該ライトデータ要求を構成するライト要求領域情報と、前記協調制御情報 28 とに基づき、前記該ライトデータ要求を自身が処理すべきか否かを判別する。

【0047】

自身が処理すると判断したディスク装置では、前記バッファ部 12 の送受信データ格納部 27 を確保し、上位装置に対しライトデータの転送開始要求を送信する。

【0048】

更に、上記ライトデータの転送開始要求を受信した上位装置 2 から送信されるライトデータを、予め確保した送受信データ格納部 27 に格納する。更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部 11 に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報を生成送信し、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【0049】

一方、自身が処理しないと判断したディスク装置では、前記アクセス要求を構成するライト要求領域情報に基づき、後述の未更新領域管理情報を更新した後、前記アクセス要求を削除することによって、ライトデータ要求処理を完了する。

【0050】

図 6 は、前記ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、

上位装置 2-a/bからのライトデータ要求が発行された場合、特に、上位装置 2-aからのライトデータ要求がディスク装置 1-aで、上位装置 2-bからのライトデータ要求がディスク装置 1-bで処理される場合の、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れを示す図である。

【 0 0 5 1 】

上述の説明の通り、上位装置 2-a/bからのライトデータ要求は、少なくともディスク装置 1-aのインタフェース制御部 1 3 を介して、ディスク装置 1-a/bのバッファ部 1 2 に格納される。

【 0 0 5 2 】

ディスク装置 1-a/bでは、受信したアクセス要求について、自身が処理すべきか否かを判別し、自身が処理すると判断したライトデータ要求の要求元上位装置に対して、ライトデータの転送開始要求を送信し、これを契機に上位装置から送信が開始されるライトデータを、送受信データ格納部 2 7 に格納する。

更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部 1 1 に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報の生成送信処理を実施する。

【 0 0 5 3 】

ここで、ディスク装置 1-aから上位装置 2-aに対して送信されるライトデータの転送開始要求、及び、ステータス情報は、前述のリードデータの転送処理と同様に、従来通りディスク装置 1-bのインタフェース制御部 1 3 を介して上位装置 2-aに転送される。

【 0 0 5 4 】

また、上位装置 2-a/bから記憶装置システム 4 に対して送信されるライトデータについては、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を、前記ディスク装置 1-a/bのインタフェース制御部 1 3 の再送判定部 2 4 で実施する。即ち、上位装置 2-aからのライトデータについては、前記ディスク装置 1-aのインタフェース制御部 1 3 の再送判定部 2 4 で、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段に転送されること無くディスク装置 1-aのバッファ部 1 2 に格納される。

【 0 0 5 5 】

また、上位装置 2-b からのライトデータについては、前記ディスク装置 1-a のインタフェース制御部 1 3 の再送判定部 2 4 で、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段への転送が実施され、更に、前記ディスク装置 1-b のインタフェース制御部 1 3 の再送判定部 2 4 で、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段に転送されること無くディスク装置 1-b のバッファ部 1 2 に格納される。

【 0 0 5 6 】

上述の通り、本実施形態では、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する個々の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を、前記複数の記憶装置で共有する手段と、共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべきと判断した記憶装置に転送する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な記憶装置システムが実現可能となる。

【 0 0 5 7 】

尚、本実施形態では、上位装置 2 からのライトデータ要求によって、一時的にディスク装置 1-a/b に格納されたデータの不一致が生じてしまう。そこで、以下に本実施形態におけるライトデータの更新処理例を説明する。

【 0 0 5 8 】

図 7 は、前述の上位装置からのライトデータ要求を自身が処理しないと判断したディスク装置において、更新処理が実施される未更新領域管理情報 3 0 の一構成例を示す図である。

【 0 0 5 9 】

未更新領域管理情報 3 0 としては、少なくとも未更新領域に関する情報を持つものであり、同図においては、上位装置から記憶装置システムに対するアクセス要求が、アクセス開始アドレス情報と、アクセスサイズ情報とから構成される場

合を例に、未更新開始アドレス情報と、未更新サイズ情報とを持ち、更に、未更新領域に関する更新時刻情報等の更新情報を含む構成としている。

【0060】

上位装置からのライトデータ要求を受信し、自身が処理しないと判断したディスク装置では、ライトデータ要求を構成するライト要求領域情報に基づき、未更新領域管理情報30を更新することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を可能としている。

【0061】

図8は、ディスク装置1-bの未更新領域管理情報30に基づき、不一致の発生した未更新領域の更新処理を実施する場合の、更新要求、及び、更新データの転送処理等の流れを示す図である。

同図において、ディスク装置1-bは、自身の未更新領域管理情報30に基づき、対を成すディスク装置1-aに対し、更新要求を発行する。該更新要求を受領したディスク装置1-aでは、必要ならばディスク部11からのディスクリードを実施し、要求対象のデータをディスク装置1-bに対し送信する。ディスク装置1-bでは、上記ディスク装置1-aから受信したデータをディスク部11に格納し、当該処理に掛かる未更新領域情報を未更新領域管理情報30から削除することによって、未更新領域の更新処理を完了する。

【0062】

尚、上述の実施形態において、上位装置からのライトデータ要求を受信し、自身が処理しないと判断したディスク装置が、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実施するものとしているが、これに限るものではなく、例えば、自身が処理すると判断したディスク装置が、ライトデータ要求を構成するライト要求領域情報に基づき、未更新領域管理情報30を更新することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実施することも可能である。

【0063】

更に、未更新領域の更新処理を開始する契機は任意である。例えば、未更新領域管理情報30に登録されたエントリー数、或いは、総未更新領域サイズ等に基づ

づき開始することが可能である。

【0064】

以下、本発明に係る第二の実施形態を、図9、10、11、12、13を用いて説明する。本実施形態では、受信したアクセス要求を、まず受信したディスク装置で処理すべきか否かを判別する。判別の結果、当該ディスク装置で処理すべきものでなければ、他のディスク装置にアクセス要求を転送する、ことに特徴がある。つまり、本実施形態では、上位装置から記憶装置システムに対するアクセス要求を、記憶装置システムを構成する一台の記憶装置で受領し、更に、当該記憶装置において、アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現可能にするものである。

【0065】

図9は、前記記憶装置システム4を構成するディスク装置1の一構成例を示す図である。同図において、上位装置2から受領したアクセス要求を、対を成すディスク装置に転送するか否かを判定制御するコマンド転送制御部29を新たに具備したことを除いては、図2と同一の構成となっている。

【0066】

また、前記協調制御情報28としては、上記上位装置2から受領したアクセス要求を、対を成すディスク装置に転送するか否かの判定制御を実施する為の情報等を保持している。更に、前記再送判定部24では、少なくとも、上位装置2から記憶装置システム4に対するフレーム等の情報の再送制御として、ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を実施するものである。

【0067】

以下、上位装置から記憶装置システムに対するアクセス要求について、アクセス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを個々のディスク装置で判別する場合のリード／ライトデータ転送処理を例に動作を説明す

る。

【0068】

図10は、前記ディスク装置1-a/bから構成される記憶装置システム4に対し、上位装置2からのリードデータ要求が発行された場合の、各装置における処理の流れを示す図である。尚、以下の説明では、ディスク装置1-aが上位装置2からのアクセス要求を受領し、該アクセス要求を処理すべき装置の判別、更に、必要な場合、前記アクセス要求をディスク装置1-bに転送に対して転送制御を実施するものとしている。

【0069】

(1)アクセス要求（コマンド）受信

上位装置2から記憶装置システム4に対して発行されたリードデータ要求は、フレーム情報として、先ず、ディスク装置1-aの前記受信部20、及び受信フレーム処理部22を介して、バッファ部12の受信コマンド格納部26に格納される。

【0070】

(2)アクセス要求（コマンド）解釈及びアクセス要求（コマンド）転送処理

ディスク装置1-aは、上記受信コマンド格納部26に格納したアクセス要求を解釈し、該アクセス要求がリードデータ要求であることを検出する。更に、該リードデータ要求を構成するリード要求領域情報と、前記協調制御情報28とに基づき、該リードデータ要求を自身が処理すべきか否かを判別する。

【0071】

上記判定で、自身が処理すると判断した場合、アクセス要求はディスク装置1-aで処理される。また、自身で処理しないと判断した場合、ディスク装置1-aは、前記上位装置2からのリードデータ要求をディスク装置1-bに対し転送する。

【0072】

ディスク装置1-bでは、ディスク装置1-aから転送されたリードデータ要求を、受信部20、及び受信フレーム処理部22を介して、バッファ部12の受信コマンド格納部26に格納する。更に、ディスク装置1-bは、受信コマンド格納部26に格納したアクセス要求を解釈し、アクセス要求がリードデータ要求である

ことを検出する。

【 0 0 7 3 】

(3) アクセス要求 (コマンド) 実行

自身が処理すると判断したディスク装置 1-a、或いは、ディスク装置 1-a からアクセス要求を転送されたディスク装置 1-b では、リードデータ要求に基づき、前記ディスク制御部 1 4 を介してディスク部 1 1 からバッファ部 1 2 の送受信データ格納部 2 7 に対するディスクリード処理を開始する。更に、前記インタフェース制御部 1 3 を介して上位装置 2 に対するリードデータ送信処理を実施する。更に、リードデータ送信後、リードデータ要求に対するステータス情報を生成送信し、アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。尚、ディスク装置 1-a において自身が処理しないと判断した場合、リードデータ要求を削除することによって、リードデータ要求に掛かる処理を完了する。

【 0 0 7 4 】

図 1 1 は、ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、上位装置 2-a/b からのリードデータ要求が発行された場合、特に、上位装置 2-a からのリードデータ要求がディスク装置 1-a で、上位装置 2-b からのリードデータ要求がディスク装置 1-b で処理される場合の、リードデータ要求、及び、リードデータの転送処理等の流れを示す図である。

【 0 0 7 5 】

上述の説明の通り、上位装置 2-a からのリードデータ要求は、ディスク装置 1-a によって受領され、ディスク装置 1-b に転送されることは無い。また、上位装置 2-b からのリードデータ要求は、上述のディスク装置 1-a の転送制御処理によって、ディスク装置 1-b が受領する。

【 0 0 7 6 】

ディスク装置 1-a/b では、受領したリードデータ要求に対して、ディスク部 1 1 からのディスクリード処理、及び、上位装置 2-a/b に対するリードデータ送信処理、更に、ステータス生成送信処理を実施する。

【 0 0 7 7 】

ここで、ディスク装置 1-a が送信するリードデータ等は、上位装置 2-a に対するものであることから、従来通りディスク装置 1-b のインタフェース制御部 1 3 を介して上位装置 2-a に転送される。

【0078】

図 1 2 は、ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、上位装置 2 からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【0079】

(1) アクセス要求 (コマンド) 受信

上位装置 2 から記憶装置システム 4 に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置 1-a の前記受信部 2 0、前記受信フレーム処理部 2 2 を介して、バッファ部 1 2 の前記受信コマンド格納部 2 6 に格納される。

【0080】

(2) アクセス要求 (コマンド) 解釈及びアクセス要求 (コマンド) 転送処理

ディスク装置 1-a は、受信コマンド格納部 2 6 に格納したアクセス要求を解釈し、アクセス要求がライトデータ要求であることを検出する。更に、該ライトデータ要求を構成するライト要求領域情報と、前記協調制御情報 2 8 とに基づき、該ライトデータ要求を自身が処理すべきか否かを判別する。

【0081】

この判定で、自身が処理すると判断した場合、前記アクセス要求はディスク装置 1-a で処理される。また、自身で処理しないと判断した場合、ディスク装置 1-a は、上位装置 2 からのライトデータ要求をディスク装置 1-b に転送する。

【0082】

ディスク装置 1-b では、ディスク装置 1-a によって転送された前記ライトデータ要求を、前記受信部 2 0、及び前記受信フレーム処理部 2 2 を介して、バッファ部 1 2 の前記受信コマンド格納部 2 6 に格納する。更に、ディスク装置 1-b は、受信コマンド格納部 2 6 に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出する。

【 0 0 8 3 】

(3) アクセス要求 (コマンド) 実行

自身が処理すると判断したディスク装置 1-a、或いは、ディスク装置 1-a からアクセス要求を転送されたディスク装置 1-b では、前記ライトデータ要求に基づき、バッファ部 12 の送受信データ格納部 27 を確保し、上位装置に対しライトデータの転送開始要求を送信する。更に、上記ライトデータの転送開始要求を受信した上位装置から送信されるライトデータを、上記予め確保した送受信データ格納部 27 に格納する。

【 0 0 8 4 】

更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部 11 に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報を生成送信する。

【 0 0 8 5 】

(4) アクセス要求 (コマンド) 完了処理

ディスク装置 1-b では、ステータス情報を生成送信の後、アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。また、ディスク装置 1-a では、上記ステータス情報を生成送信の後、或いは、自身で処理しないと判断し、上位装置 2 からのライトデータ要求をディスク装置 1-b に対し転送した後、ライトデータ要求を構成するライト要求領域情報に基づき、後述の未更新領域管理情報を更新する。更に、アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【 0 0 8 6 】

本実施形態において、前記ディスク装置 1-a/b から構成される記憶装置システム 4 に対し、上位装置 2-a/b からのライトデータ要求が発行された場合、特に、上位装置 2-a からのライトデータ要求がディスク装置 1-a で、上位装置 2-b からのライトデータ要求がディスク装置 1-b で処理される場合の、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れは、図 6 と同一であることから省略する。

【 0 0 8 7 】

但し、上述の説明の通り、上位装置 2-a/bからのアクセス要求は、少なくともディスク装置 1-aのバッファ部 12に格納され、ディスク装置 1-aが自身で処理しないと判断した場合に、ディスク装置 1-bに対して転送されるものである。

【 0 0 8 8 】

上述の通り、本実施形態では、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を受領する。更に、当該記憶装置において、アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送する手段を具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な記憶装置システムが実現可能となる。

【 0 0 8 9 】

また、前述の実施形態と同様に、本実施形態でも上位装置 2からのライトデータ要求によって、一時的にディスク装置 1-a/bに格納されたデータの不一致が生じてしまう。そこで、以下に本実施形態におけるライトデータの更新処理例を説明する。

【 0 0 9 0 】

本実施形態では、前述の未更新領域管理情報 30に対して、新たに自身の格納するデータが最新か否かの情報（以下、更新種別情報）を付加した未更新領域管理情報をディスク装置 1-aで保持し、ディスク装置 1-aにおいて、上位装置からのライトデータ要求を構成するライト要求領域情報に基づき、上記未更新領域管理情報を構成する未更新領域情報と、更新種別情報とを更新・管理することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実現するものである。

【 0 0 9 1 】

図 13は、前記ディスク装置 1-aの未更新領域管理情報に基づき、不一致の発生した未更新領域の更新処理を実施する場合の、更新要求、及び、更新データの転送処理等の流れを示す図である。また、同図において、ディスク装置 1-aがデ

ディスク装置 1-b に格納された最新のデータを用いて、自身を更新する場合を例としている。

【 0 0 9 2 】

ディスク装置 1-a は、少なくとも、未更新領域情報と更新種別情報とから構成される前記未更新領域管理情報に基づき、対を成すディスク装置 1-b に対し、更新要求（最新データ送信要求）を発行する。該更新要求を受領したディスク装置 1-b では、必要ならば前記ディスク部 1 1 からのディスクリードを実施し、要求対象のデータをディスク装置 1-a に対し送信する。ディスク装置 1-a では、上記ディスク装置 1-b から受信したデータをディスク部 1 1 に格納し、当該処理に掛かる未更新領域情報を前記未更新領域管理情報 3 0 から削除することによって、未更新領域の更新処理を完了する。

【 0 0 9 3 】

また、ディスク装置 1-a が自身の格納する最新データを用いて、ディスク装置 1-b を更新する場合、ディスク装置 1-a は、前記未更新領域管理情報に基づき、対を成すディスク装置 1-b に対し、更新要求（最新データ受信要求）を発行し、引き続き更新対象のデータを送信する。更新要求及び更新対象データを受領したディスク装置 1-b では、上記ディスク装置 1-a から受信したデータをディスク部 1 1 に格納する。更に、ディスク装置 1-a は、更新対象データの送信後、当該処理に掛かる未更新領域情報を前記未更新領域管理情報 3 0 から削除することによって、未更新領域の更新処理を完了する。

【 0 0 9 4 】

以上の実施形態では、上位装置から記憶装置システムに対するアクセス要求を構成するアクセス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを判別する場合を例に動作を説明してきたが、これに限定するものではない。

【 0 0 9 5 】

例えば、記憶装置がオブジェクト（ファイル）単位で格納するデータを管理する場合、上位装置から記憶装置システムに対するアクセス要求を構成するオブジェクト（ファイル）情報を条件として、該アクセス要求を自身が処理すべきか否

かを判別することが可能であり、前記未更新領域管理情報として少なくとも保持する未更新領域情報も、オブジェクト（ファイル）情報を用いることで、上述の未更新領域の更新処理を実現することが可能となる。

【 0 0 9 6 】

また、以上の実施形態では、複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、当該記憶装置システムを構成する個々の記憶装置が、或いは、少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求について、自身が処理すべきか否かを判別する手段を具備することによって、個々の記憶装置を制御する制御装置を不要とする。また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現するものである。

【 0 0 9 7 】

しかし、最適な記憶装置の利用効率を実現する為、上位装置からのライトデータ要求によって、一時的にはあるが、二重化されたディスク装置のデータに不一致が生じ、上記ライトデータの更新処理を必要とする場合もある。また、上記上記ライトデータの更新処理において、前記インタフェース3を介したライトデータの転送処理を実施する為、インタフェース3の負荷を軽減するものではない。

【 0 0 9 8 】

そこで、上位装置2から記憶装置システム4に対するアクセス要求、特に、ライトデータ要求を、該記憶装置システム4を構成する複数の記憶装置1で共有し、更に、対を成す複数の記憶装置が同期して該ライトデータ要求を処理することによって、インタフェース負荷を軽減することによる高性能化を図った実施形態について、以下に説明する。

【 0 0 9 9 】

以下、本発明に係る第三の実施形態を、図14、15を用いて説明する。

本実施形態では、上位装置2から記憶装置システム4に対するライトデータ要求を、該記憶装置システム4を構成する複数の記憶装置1で共有し、更に、対を成す複数の記憶装置でライトデータの転送処理をも同期して実施することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、インタフェー

ス負荷を軽減する高性能な前記記憶装置システムを実現するものである。

【0100】

尚、本実施形態の記憶装置システムを構成する記憶装置（ディスク装置）の構成は、再送判定制御部24において、上位装置2から記憶装置システム4に対するフレーム等の情報の再送制御として、ライトデータについても、前記協調制御情報28に基づく再送制御を実施することを除いて、図2と同一であることから省略する。

【0101】

また、上位装置から記憶装置システムに対するリードデータ要求については、本実施形態の対象外であり、上述の実施形態の何れかを採用するものとして、その説明も省略する。

【0102】

以下、上位装置から記憶装置システムに対するライトデータ要求について、対を成す二台の記憶装置でライトデータの転送処理を同期して実施する場合のライトデータ転送処理を例に動作を説明する。

図14は、前記ディスク装置1-a/bから構成される記憶装置システム4に対し、上位装置2からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【0103】

図15は、前記ディスク装置1-a/bから構成される記憶装置システム4に対し、上位装置2からのライトデータ要求が発行された場合の、ライトデータ要求、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れを示す図である。

【0104】

(1) アクセス要求（コマンド）受信及び再送処理

上位装置2から記憶装置システム4に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置1-aの前記受信部20を介して受信される。

ディスク装置1-aでは、前記ライトデータ要求を、前記受信フレーム処理部2

2 を介して、前記バッファ部 1 2 の前記受信コマンド格納部 2 6 に格納する。

【0105】

また、ディスク装置 1-a の前記再送判定部 2 4 は、前記協調制御情報 2 8 に基づき、上位装置から記憶装置システム 4 に対するアクセス要求を後段のディスク装置 1-b に対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部 2 5 に対する再送制御（指示）信号を生成する。結果、前記出力選択部 2 5 は、該再送制御（指示）信号に基づき、前記上位装置からのライトデータ要求を後段のディスク装置 1-b に対し再送する。

【0106】

ディスク装置 1-b では、前記ディスク装置 1-a によって再送された前記ライトデータ要求を、前記受信フレーム処理部 2 2 を介して、前記バッファ部 1 2 の前記受信コマンド格納部 2 6 に格納する。また、ディスク装置 1-b の前記再送判定部 2 4 は、前記協調制御情報 2 8 に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【0107】

(2) アクセス要求（コマンド）解釈及びアクセス要求（コマンド）実行

ディスク装置 1-a/b は、前記受信コマンド格納部 2 6 に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出し、バッファ部 1 2 の送受信データ格納部 2 7 を確保する。ディスク装置 1-a では、自身の送受信データ格納部 2 7 を確保したことを契機として、ディスク装置 1-b に対しライトデータの転送開始要求を送信する。一方、ディスク装置 1-b では、自身の送受信データ格納部 2 7 を確保したことと、ディスク装置 1-a からのライトデータの転送開始要求受信を契機として、上位装置 2 に対しライトデータの転送開始要求を送信する。

【0108】

上位装置 2 では、上記ライトデータの転送開始要求を受信したことを契機として、ライトデータの送信を開始する。

上位装置 2 から送信されたライトデータは、先ず、ディスク装置 1-a の前記受

信部 2 0 を介して受信される。ディスク装置 1 -a では、上記ライトデータを、前記受信フレーム処理部 2 2 を介して、バッファ部 1 2 の前記送受信データ格納部 2 7 に格納する。

【 0 1 0 9 】

また、ディスク装置 1 -a の再送判定部 2 4 は、協調制御情報 2 8 に基づき、上位装置 2 から記憶装置システム 4 に対するライトデータを後段のディスク装置 1 -b に対して再送するように設定しておくことによって、受信したフレーム情報がライトデータであることを検出し、出力選択部 2 5 に対する再送制御（指示）信号を生成する。出力選択部 2 5 は、再送制御（指示）信号に基づき、上位装置からのライトデータを後段のディスク装置 1 -b に対し再送する。

【 0 1 1 0 】

ディスク装置 1 -b では、ディスク装置 1 -a によって再送されたライトデータを、受信フレーム処理部 2 2 を介して、バッファ部 1 2 の前記送受信データ格納部 2 7 に格納する。また、ディスク装置 1 -b の再送判定部 2 4 は、協調制御情報 2 8 に基づき、ライトデータを後段の装置に対し再送しないよう設定されている。

【 0 1 1 1 】

ディスク装置 1 -a では、ライトデータ受信後、或いは、受信したライトデータをディスク部 1 1 に対しライトするディスクライト処理完了後、ディスク装置 1 -b に対し、前記ライトデータ要求に対するステータス情報を生成送信する。更に、当該ライトデータ要求を削除することによって、一連のライトデータ要求処理を完了する。

【 0 1 1 2 】

ディスク装置 1 -b では、ライトデータ受信後、或いは、受信したライトデータをディスク部 1 1 に対しライトするディスクライト処理完了後、ディスク装置 1 -a からのステータス情報を受信後、上位装置 2 に対して当該ライトデータ要求に対するステータス情報を生成送信し、更に、当該ライトデータ要求を削除することによって、一連のライトデータ要求処理を完了する。

【 0 1 1 3 】

上述の通り、本実施形態では、複数の記憶装置から構成される RAID1 の記憶装

置システムにおいて、該記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対を成す複数の記憶装置間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、記憶装置間でのデータ更新処理を不要とし、インタフェース負荷を軽減する高性能な前記記憶装置システムが実現可能となる。

【0114】

以上の実施形態では、本発明を適用した複数の記憶装置のみから構成される、RAID1の記憶装置システムについて述べてきたが、これに限定されるものではない。例えば、従来のディスクアレイ装置、即ち、複数のディスク装置と、これらのディスク装置を制御する制御装置とから構成されるディスクアレイ装置においても、本発明を適用したディスク装置を採用することによって、上記制御装置の構成が容易となり、結果、低価格化を実現することが可能となる。

【0115】

更に、前述の上位装置からのアクセス要求を自身が処理すべきか否かの判別処理の一例を、以下に説明する。

【0116】

図16は、前記協調制御情報28を構成する情報の一例を示す図である。

同図において、協調制御情報28を構成する情報として、前記記憶装置システム4の装置識別情報と、個々の記憶装置自身に固有の装置識別情報と、自身と協調制御を実施する記憶装置に固有の装置識別情報と、協調処理モード情報と、更に、協調処理対象となる領域管理情報として、協調領域サイズ情報と、アクティブ領域開始アドレス情報と、アクティブ領域サイズ情報とを保持するものである。

【0117】

また、上記各情報について、記憶装置1-a/bにおける設定例を示しており、前記記憶装置システム4の装置識別情報は、記憶装置1-a/bともにID_0、個々の記

憶装置自身に固有の装置識別情報は、記憶装置 1-a に ID_1、記憶装置 1-b に ID_2、自身と協調制御を実施する記憶装置に固有の装置識別情報は、記憶装置 1-a に ID_2、記憶装置 1-b に ID_1 を設定し、協調処理モード情報は、記憶装置 1-a にコマンド再送、記憶装置 1-b にコマンド受領を設定し、更に、協調領域サイズ情報は N として、アクティブ領域開始アドレス情報は、記憶装置 1-a に 0、記憶装置 1-b に $N/2$ 、アクティブ領域サイズ情報は、記憶装置 1-a/b とともに $N/2$ を設定したものである。

【 0 1 1 8 】

尚、上記記憶装置システム 4 の装置識別情報と、個々の記憶装置自身に固有の装置識別情報と、自身と協調制御を実施する記憶装置に固有の装置識別情報と、協調処理モード情報とは、後述のフレームデータ再送処理判定で用いられる協調制御情報である。

【 0 1 1 9 】

図 1 7 は、前記上位装置からのアクセス要求情報として、FCP (Fibre Channel Protocol) で規定されるアクセス要求情報の構成例を示す図である。同図において、上記アクセス要求情報は、アクセス要求対象領域の開始アドレス情報 (図中、Logical Block Address) と、アクセス要求対象領域サイズ情報 (Transfer Length) を含むものである。

【 0 1 2 0 】

上位装置からのアクセス要求情報を受領した記憶装置 1 は、自身の保持する前記協調制御情報の、アクティブ領域開始アドレス情報と、アクティブ領域サイズ情報と、上記上位装置からのアクセス要求情報の、アクセス要求対象領域の開始アドレス情報と、アクセス要求対象領域サイズ情報とに基づき、該アクセス要求を自身が処理すべきか否か、即ち、自身のアクティブ領域に対するアクセス要求か否かを判別し、自身のアクティブ領域に対するアクセス要求についてのみ、処理を実施する。

【 0 1 2 1 】

例えば、上記上位装置からのアクセス要求対象領域が、 $N/2$ 以下の場合、記憶装置 # 1 が、また、 $N/2$ 以上の場合、記憶装置 # 2 が、該上位装置からのア

クセス要求を処理する。更に、上記上位装置からのアクセス要求対象領域が、記憶装置 # 1 及び記憶装置 # 2 のアクティブ領域に係る場合、記憶装置 # 1 及び記憶装置 # 2 が、夫々自身のアクティブ領域に対するアクセス要求を処理する。

【 0 1 2 2 】

また、上述の実施形態では、協調処理対象領域の管理情報として、アドレス情報と、サイズ情報とを保持し、上位装置からのアクセス要求情報を構成するアドレス情報と、サイズ情報とに基づき、該アクセス要求を自身が処理すべきか否かを判別する場合を例に説明したが、これに限るものではなく、例えば、図 1 7 に示す論理ユニット番号（図中、FCP_LUN or Logical Unit Number）を条件とすることも可能であり、この場合、前記記憶装置 1 は、前記協調制御情報 2 8 を構成する協調処理対象領域に関する管理情報として、アクティブ論理ユニット番号情報を保持するものである。

【 0 1 2 3 】

更にまた、上位装置からのアクセス要求情報が、ファイル名等のオブジェクト名を保持する場合、前記記憶装置 1 は、前記協調制御情報 2 8 を構成する協調処理対象領域に関する管理情報として、アクティブオブジェクトリスト情報を保持することによって、オブジェクト単位で、上記上位装置からのアクセス要求を自身が処理すべきか否かを判別することも可能である。

【 0 1 2 4 】

尚、前述の第二の実施形態においては、上位装置からのアクセス要求対象領域が、ディスク装置 1-a 及び前記ディスク装置 1-b のアクティブ領域に係る場合、ディスク装置 1-a が、上位装置からのアクセス要求情報を、ディスク装置 1-b のアクティブ領域に対するアクセス要求に加工して前記ディスク装置 1-b に対して送信することによって、ディスク装置 1-a/b での協調制御処理を実現することが可能である。

【 0 1 2 5 】

或いは、前記ディスク装置 1-b において、前述の協調制御情報（アクティブ領域開始アドレス情報、アクティブ領域サイズ情報、等）を保持し、前記上位装置からのアクセス要求情報の、アクセス要求対象領域情報とに基づき、該アクセス

要求を自身が処理すべきか否か、即ち、自身のアクティブ領域に対するアクセス要求か否かを判別し、自身のアクティブ領域に対するアクセス要求についてのみ処理を実施することによって、上述の前記ディスク装置 1-a が、前記上位装置からのアクセス要求情報を、ディスク装置 1-b のアクティブ領域に対するアクセス要求に加工してディスク装置 1-b に対して送信することなく、前記ディスク装置 1-a/b での協調制御処理を実現することが可能である。

【 0 1 2 6 】

更に、前述の上位装置と記憶装置システムとの間で転送されるフレームデータ等の情報を後段の装置に対して再送するか否かの判定処理の一例を、以下に説明する。

【 0 1 2 7 】

図 1 8 は、上位装置と記憶装置システムとの間で転送される情報として、FC-PH (Fibre Channel PHYSICAL AND SIGNALING INTERFACE) で規定されるフレームヘッダ情報の構成例を示す図である。

【 0 1 2 8 】

同図において、上記フレームヘッダ情報は、フレームデータの送信先／送信元の情報 (D_ID: Destination_Identifier、S_ID: Source_Identifier) と、フレームデータ種別情報 (R_CTL: Routing Control、TYPE: Data structure type) と、フレームデータの関連するイクスチェンジ識別情報 (RX_ID: Responder Exchange_Identifier、OX_ID: Originator Exchange_Identifier) のフィールドを具備している。

【 0 1 2 9 】

先ず、前記記憶装置システム 4 (前記ディスク装置 1-a/b) から上位装置に対して送信されるフレームデータ (リードデータ、ライトデータの転送開始要求、ステータス情報、等) の場合、上記フレームヘッダ情報のフレームデータの送信先情報 (D_ID) として、上位装置の識別情報が設定されており、当該フレームデータを受信したディスク装置は、再送する。

【 0 1 3 0 】

そこで、以下では、上位装置から前記記憶装置システム 4 に対して送信される

、即ち、前記フレームヘッダ情報のフレームデータの送信先情報 (D_ID) として、記憶装置システム 4 の識別情報が設定されたフレームデータ (アクセス要求、ライトデータ) を対象に、再送判定処理の一例を説明する。

【 0 1 3 1 】

(1) アクセス要求

再送判定部 2 4 では、上位装置から記憶装置システム 4 に対するフレームデータ (アクセス要求) を受信した際、前記フレームヘッダ情報のフレームデータ種別情報 (R_CTL、TYPE) から、当該フレームデータがアクセス要求であることを検出し、更に、前記協調制御情報 2 8 を構成する協調処理モード設定情報 (コマンド再送、且つ或いは、コマンド受領) に基づき、該受信したアクセス要求を後段の装置に対して再送するか否かを決定することで、フレームデータ (アクセス要求) の再送処理を実現する。

【 0 1 3 2 】

(2) ライトデータ

再送判定部 2 4 では、上位装置から記憶装置システム 4 に対するフレームデータ (ライトデータ) を受信した際、前記フレームヘッダ情報のフレームデータ種別情報 (R_CTL、TYPE) から、当該フレームデータがライトデータであることを検出し、更に、前記フレームヘッダ情報のフレームデータの送信元情報 (S_ID) と、イクスチェンジ識別情報 (RX_ID、OX_ID) とが、後述のライト処理管理情報に登録されたものと一致するか否かを検出し、一致する場合、該ライトデータを受領し、不一致の場合、後段の装置に対して再送することで、フレームデータ (ライトデータ) の再送処理を実現する。

【 0 1 3 3 】

図 1 9 は、上記ライト処理管理情報の一構成例を示すものであり、実行中 (処理未完) のライト処理の前記イクスチェンジ識別情報 (RX_ID、OX_ID) と、アクセス要求発行元上位装置の識別情報 (Host ID) とを関連付けたものを、ライト処理管理情報として保持する例である。同図において、例えば、同一の上位装置 (Host ID (S_ID) = ID_3) から、前記記憶装置システム 4 に対して送信されるライトデータフレームは、該フレームヘッダを構成する前記イクスチェンジ識別

情報 (RX_ID and/or OX_ID) が、上記ライト処理管理情報に登録されたものと一致するか否かを検出することで、自身が受領すべきライトデータフレームか否かの検出が可能となる。

【 0 1 3 4 】

尚、上記ライト処理管理情報の最少の構成例としては、実行中（処理未完）のライト処理の前記イクスチェンジ識別情報 (OX_ID) と、アクセス要求発行元上位装置の識別情報 (Host ID) とから構成することも可能であるが、上記 OX_ID は上位装置側で設定する値であり、その独立性を保障されるものではない。これに対し、前記 RX_ID は記憶装置側で任意の値を設定可能であることから、前記ライト処理管理情報として保持することが望ましく、更に、協調制御の対象となる記憶装置間で異なる値を取るように設定することが望ましい。

【 0 1 3 5 】

以上説明した各実施形態には、以下の効果がある。

本発明の第一の実施形態によれば、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する個々の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を、前記複数の記憶装置で共有する手段と、上記共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべきと判断した記憶装置に転送する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で、また、記憶装置の利用効率の向上が期待される記憶装置システムが実現可能となる。

【 0 1 3 6 】

本発明の第二の実施形態によれば、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を受領し、更に、当該記憶装置において、上記アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、前記上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送する手段を具備することによって、個々の記憶装置を制御する制御装置が不要で、また、記憶装置の利用効率の向上が期待される記憶装置

システムが実現可能となる。

【0137】

本発明の第三の実施形態によれば、複数の記憶装置から構成される記憶装置システムにおいて、記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対を成す複数の記憶装置間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要でインタフェース負荷を軽減する記憶装置システムが実現可能となる。

【0138】

【発明の効果】

本発明により、複数の記憶装置から構成されるRAID1を含む記憶装置システムにおいて、上位装置からのアクセス要求を、記憶装置システムを構成する複数の記憶装置間で協調して処理することが可能となる。

【図面の簡単な説明】

【図1】 本発明の情報処理システムの一構成例を示す図

【図2】 本発明の第一実施形態に係る記憶装置の一構成例を示す図

【図3】 本発明の第一実施形態に係るリード処理フローの一例を示す図

【図4】 本発明の第一実施形態に係るリード処理における情報転送の一例を示す図

【図5】 本発明の第一実施形態に係るライト処理フローの一例を示す図

【図6】 本発明の第一実施形態に係るライト処理における情報転送の一例を示す図

【図7】 本発明の第一実施形態に係る未更新領域管理情報の一構成例を示す図

【図8】 本発明の第一実施形態に係る未更新領域の更新処理における情報転送の一例を示す図

【図9】 本発明の第二実施形態に係る記憶装置の一構成例を示す図

【図10】 本発明の第二実施形態に係るリード処理フローの一例を示す図

【図 1 1】 本発明の第二実施形態に係るリード処理における情報転送の一例を示す図

【図 1 2】 本発明の第二実施形態に係るライト処理フローの一例を示す図

【図 1 3】 本発明の第二実施形態に係る未更新領域の更新処理における情報転送の一例を示す図

【図 1 4】 本発明の第三実施形態に係るライト処理フローの一例を示す図

【図 1 5】 本発明の第三実施形態に係るライト処理における情報転送の一例を示す図

【図 1 6】 本発明の協調制御情報の一構成例を示す図

【図 1 7】 アクセス要求情報の一構成例を示す図

【図 1 8】 フレームヘッダ情報の一構成例を示す図

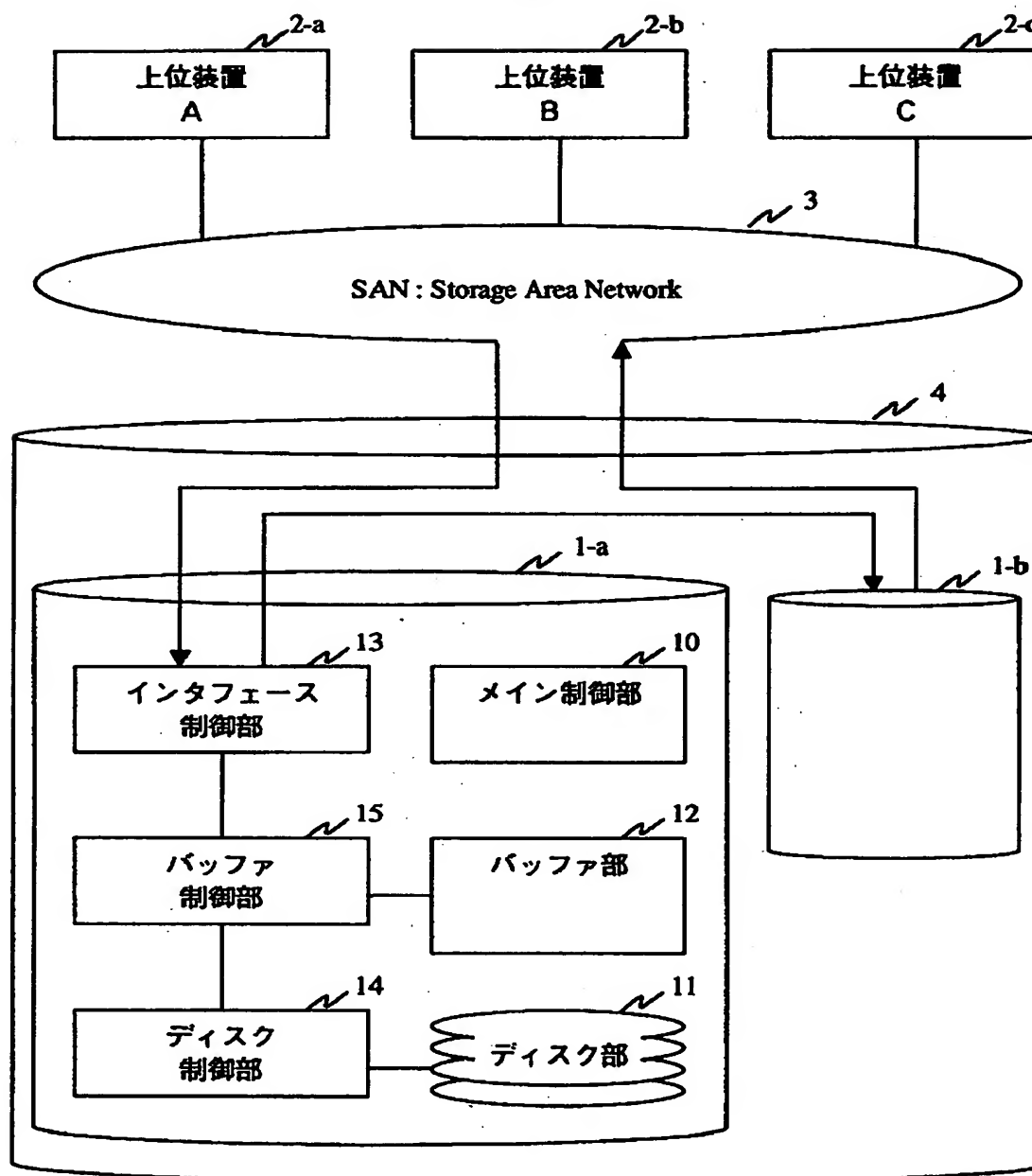
【図 1 9】 本発明のライト処理管理情報の一構成例を示す図

【符号の説明】

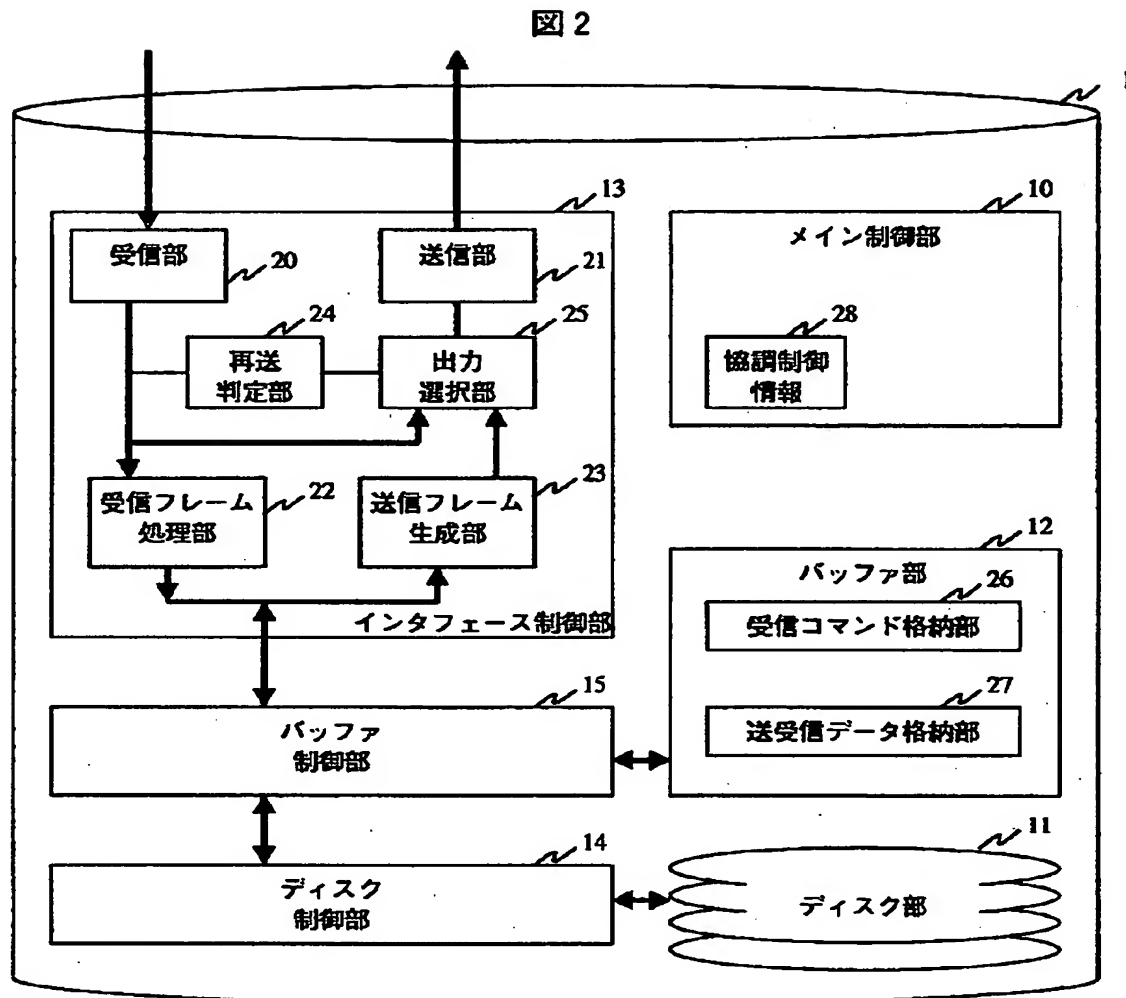
1・・・記憶装置、2・・・上位装置、3・・・インタフェース (SAN: Storage Area Network)、4・・・記憶装置システム、10・・・メイン制御部、11・・・ディスク部 (記憶媒体)、12・・・バッファ部、13・・・インタフェース制御部、14・・・ディスク制御部、15・・・バッファ制御部、20・・・受信部、21・・・送信部、22・・・受信フレーム処理部、23・・・送信フレーム生成部、24・・・再送判定部、25・・・出力選択部、26・・・受信コマンド格納部、27・・・送受信データ格納部、28・・・協調制御情報、29・・・コマンド転送制御部、30・・・未更新領域管理情報

【図 1】

圖 1

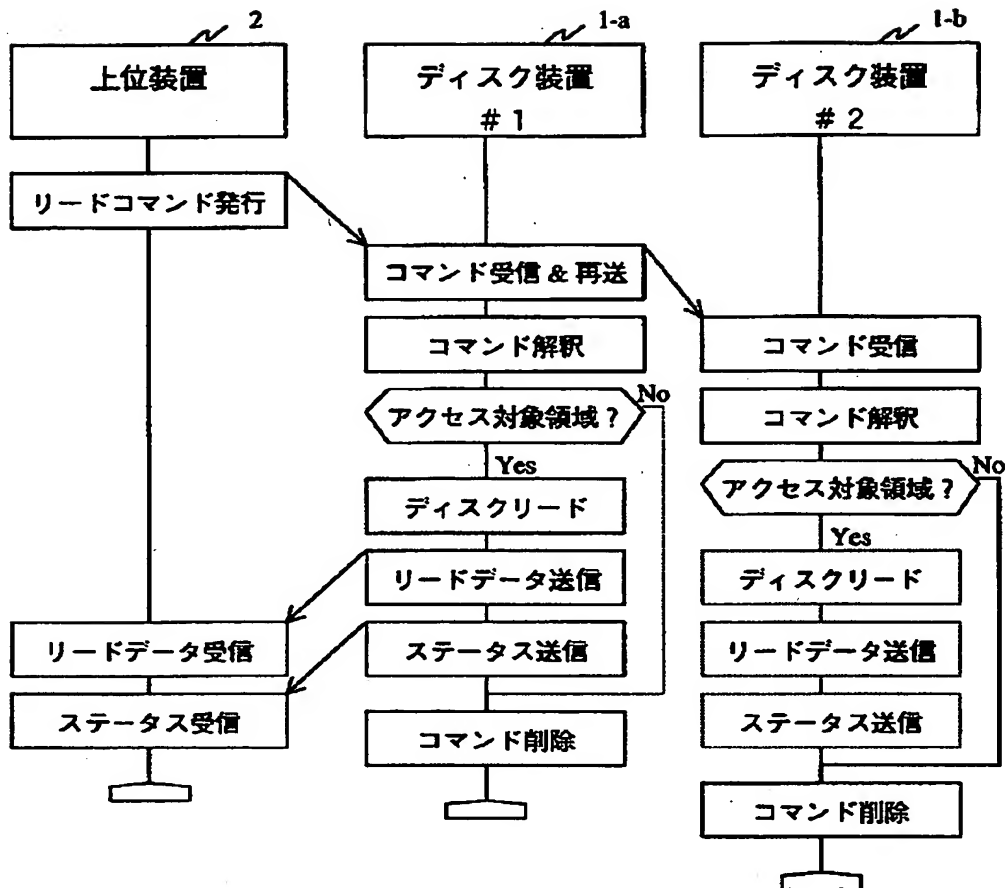


【図2】

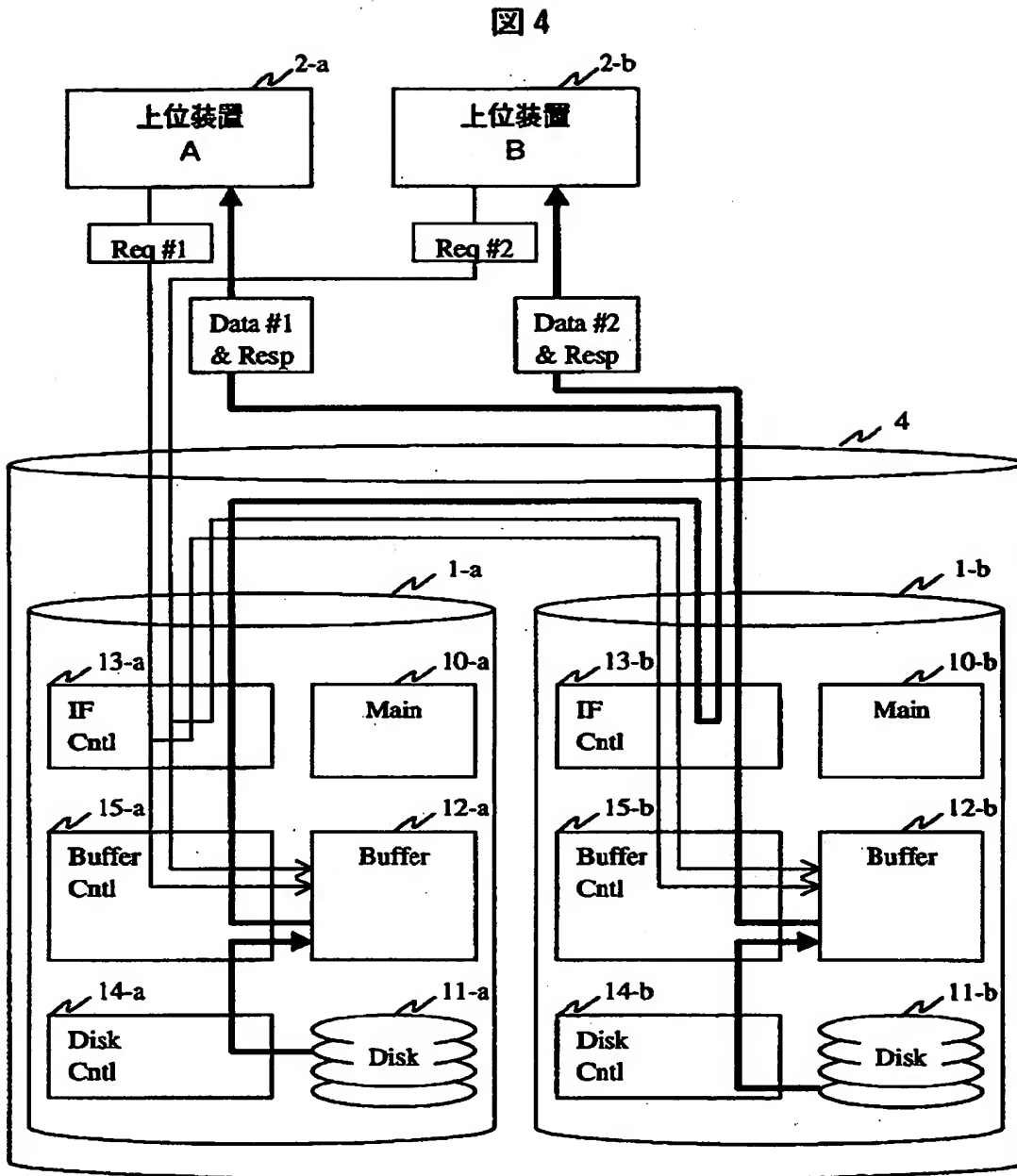


【図 3】

图 3

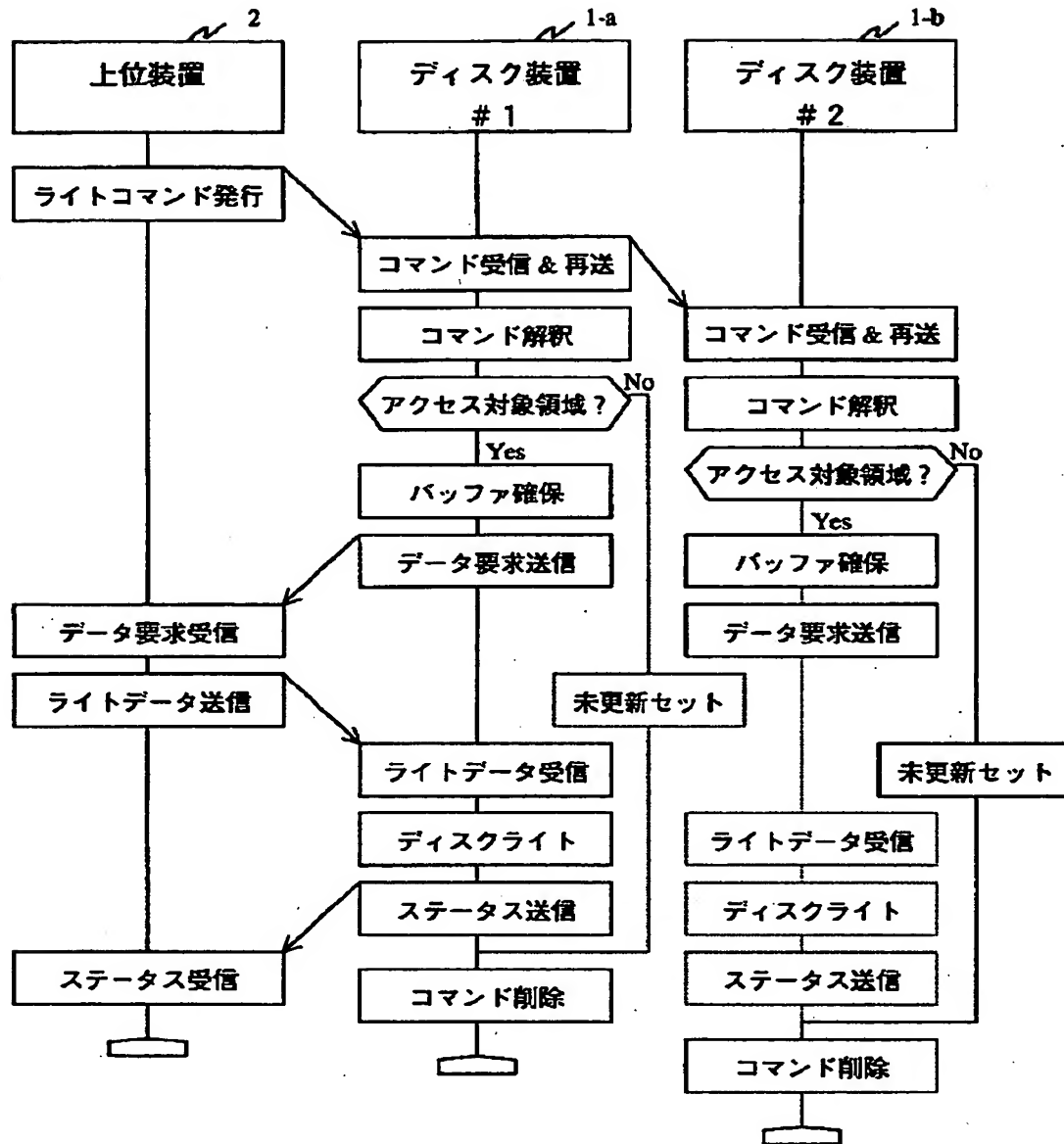


【図 4】



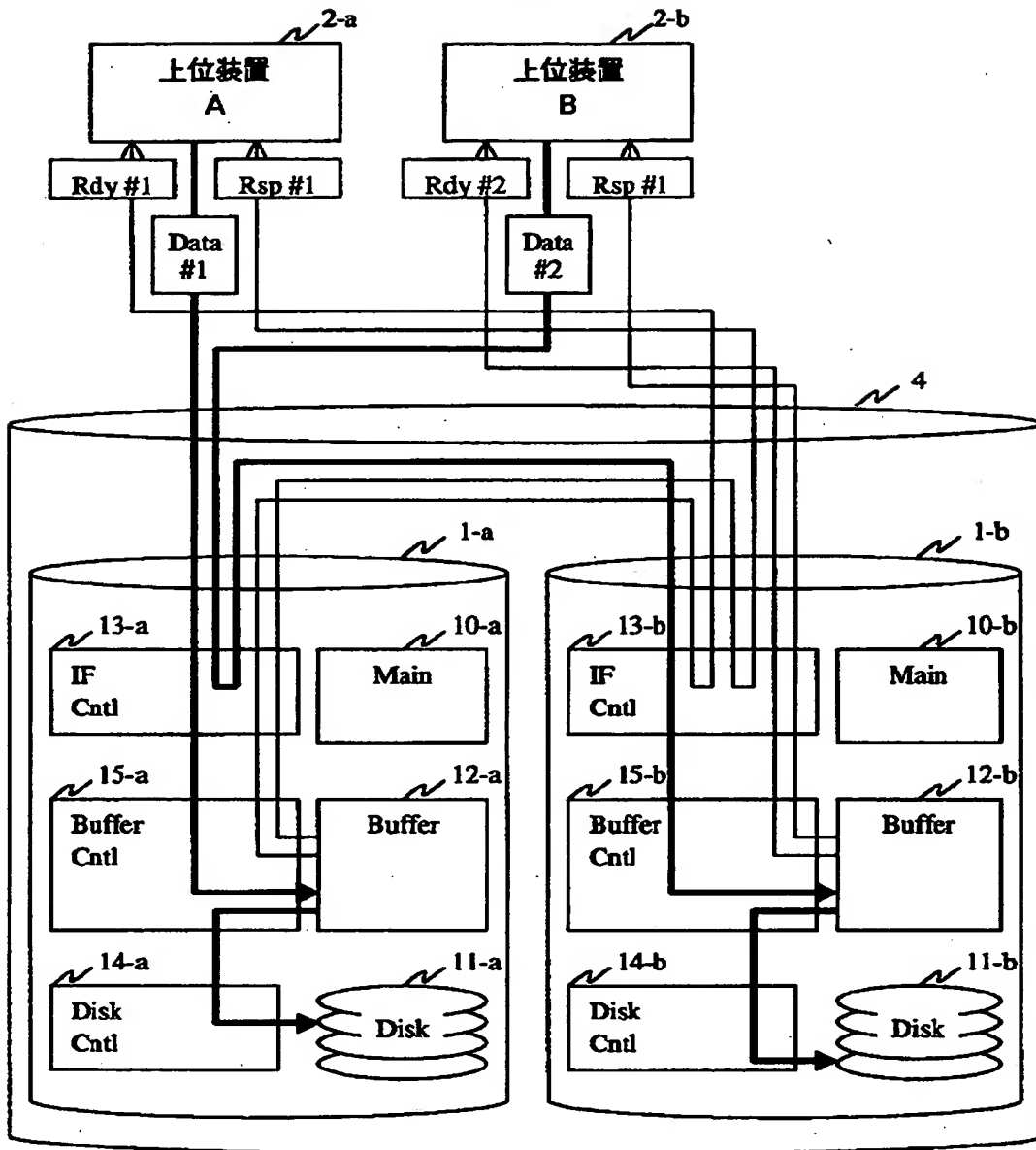
【図 5】

図 5



【図 6】

図 6



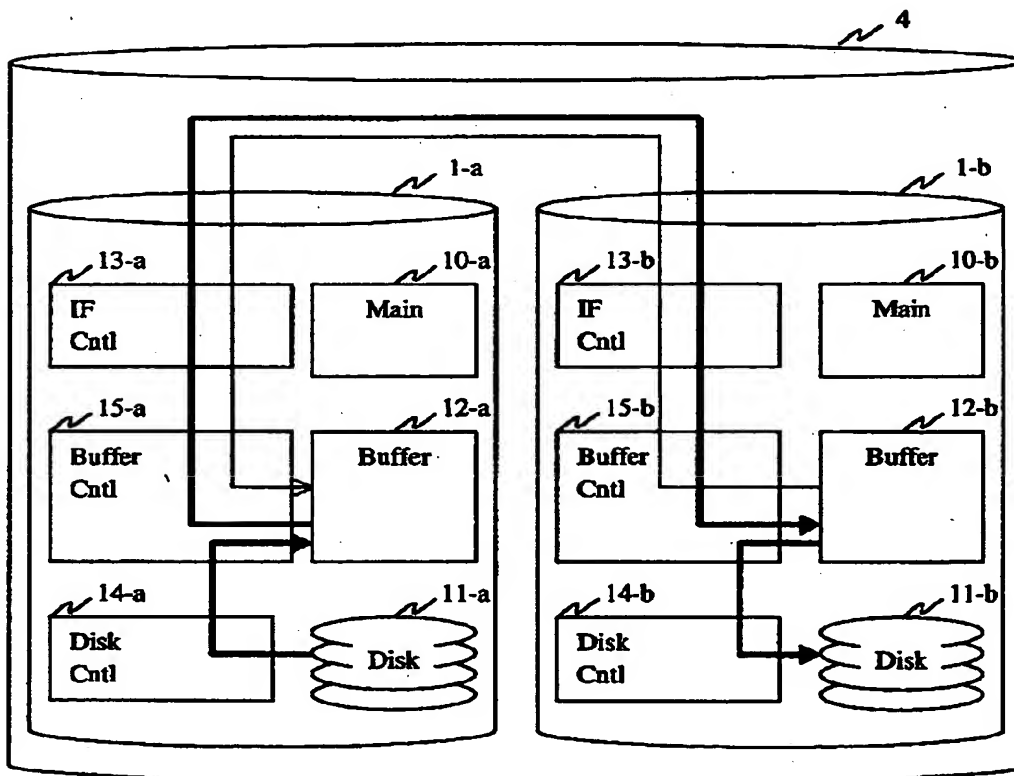
【図 7】

図 7

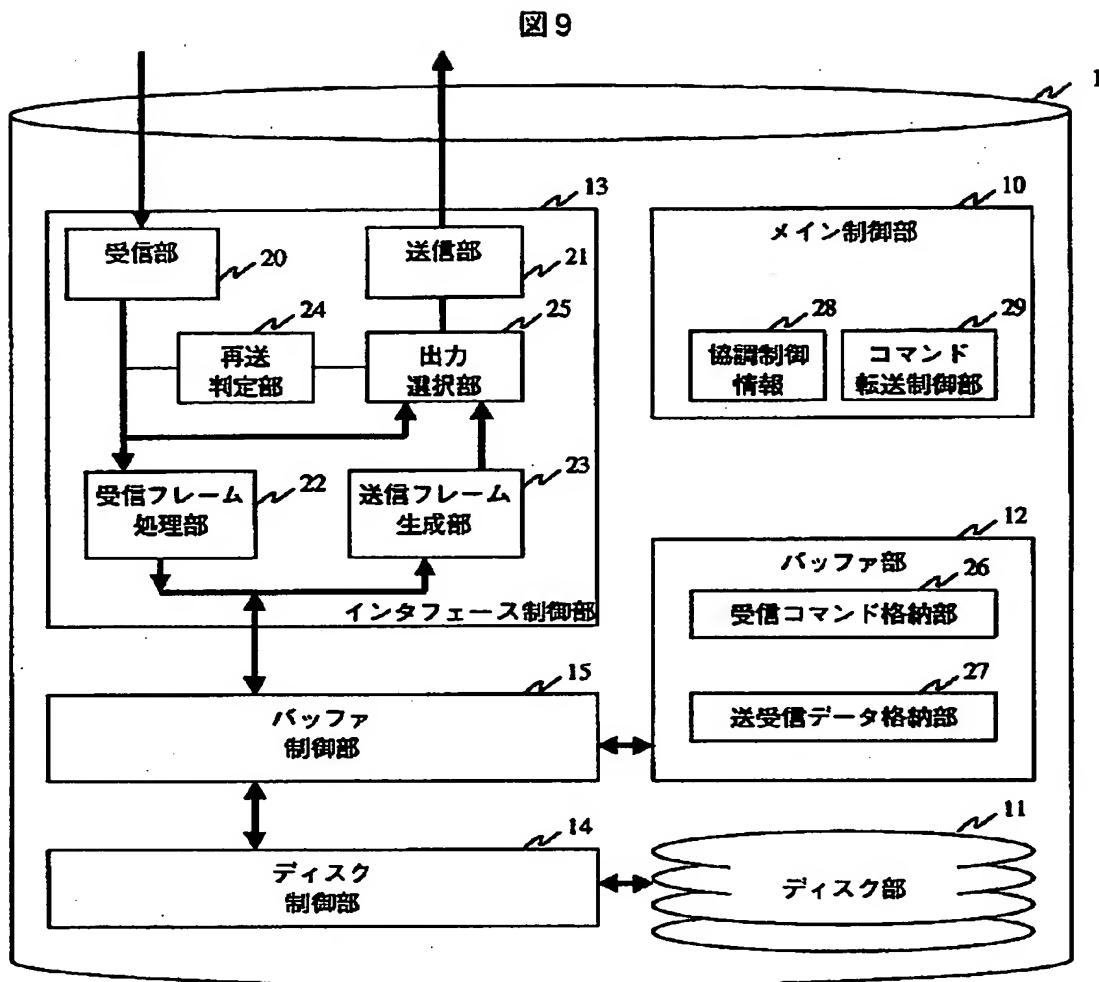
#	未更新開始アドレス	未更新サイズ	更新情報
0			
1			
2			
.			
N			

【図 8】

図 8

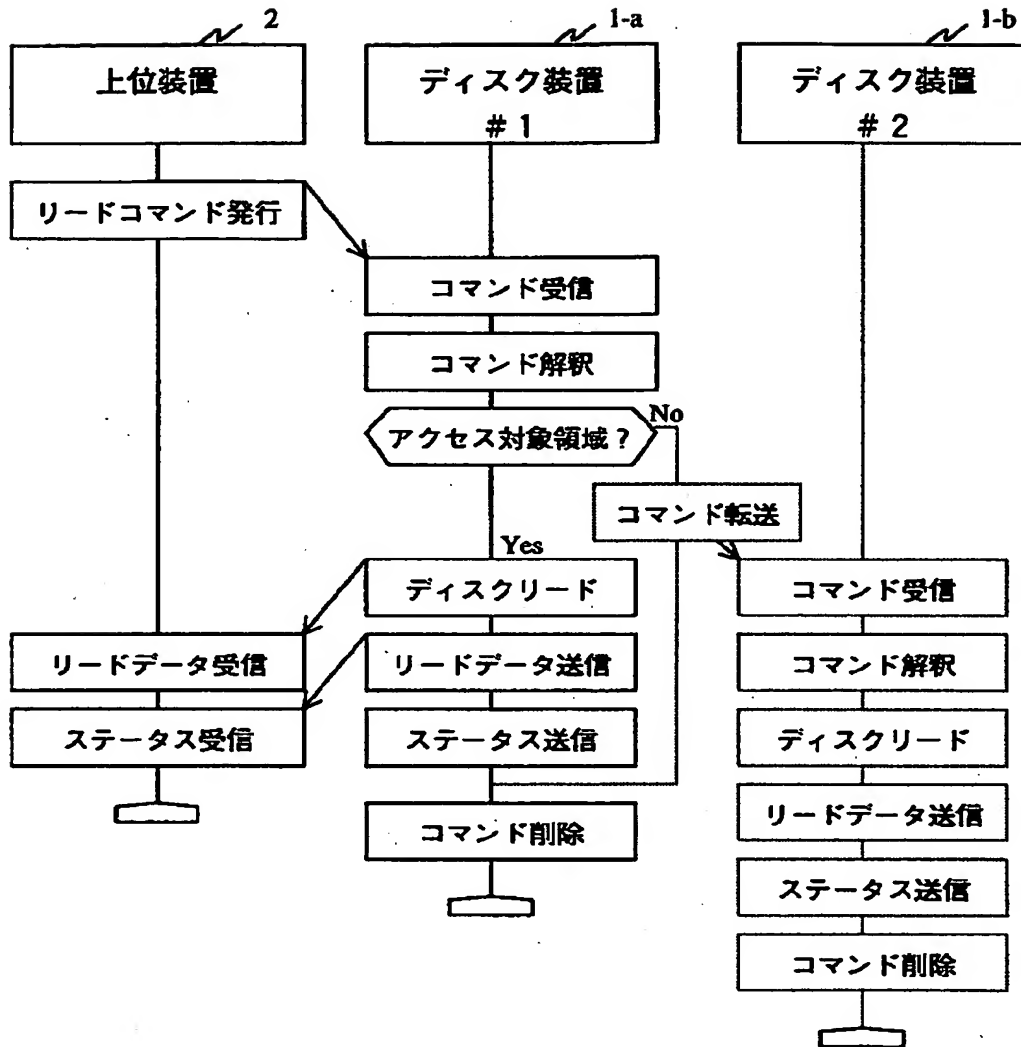


【図9】

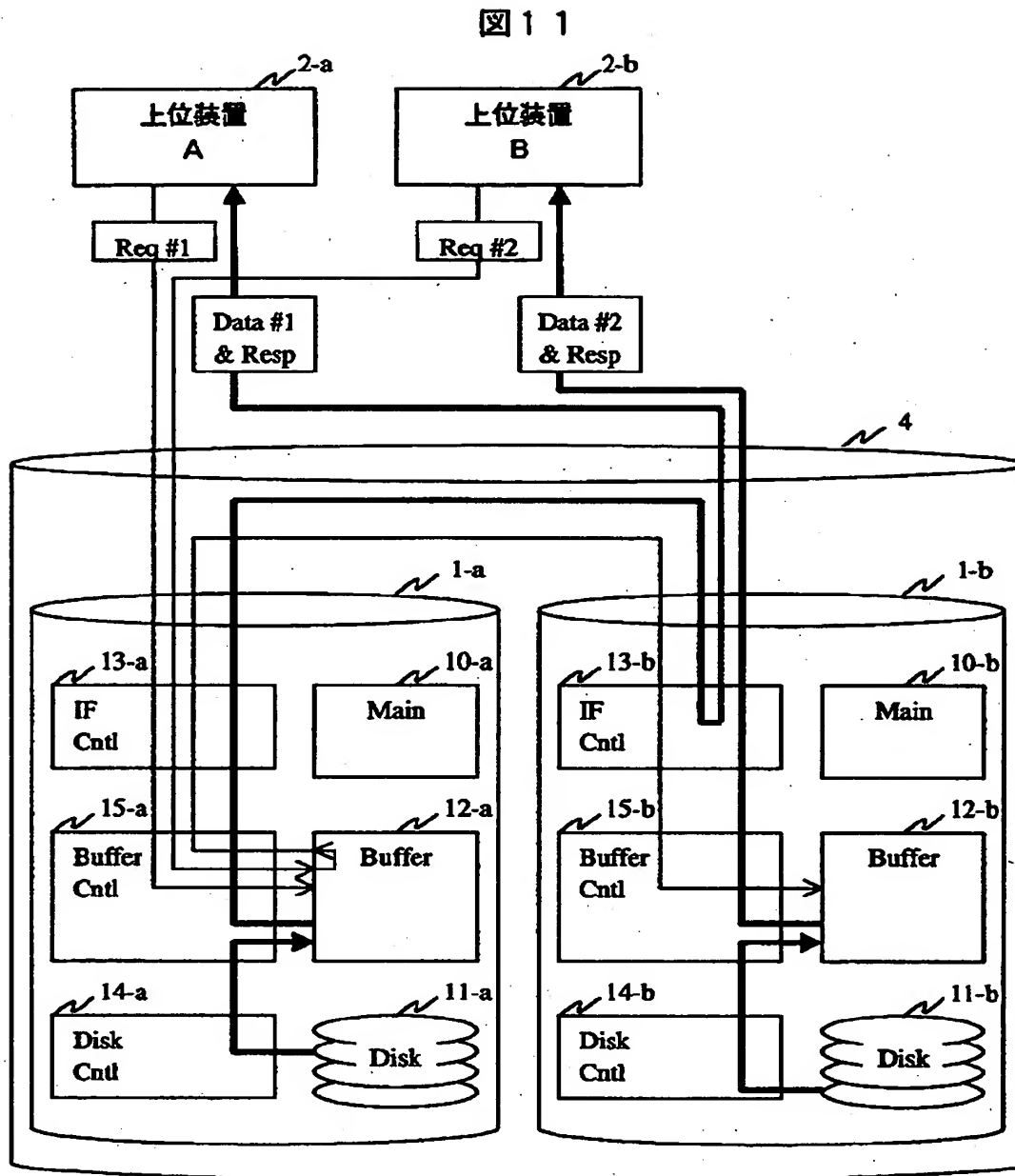


【図10】

図10

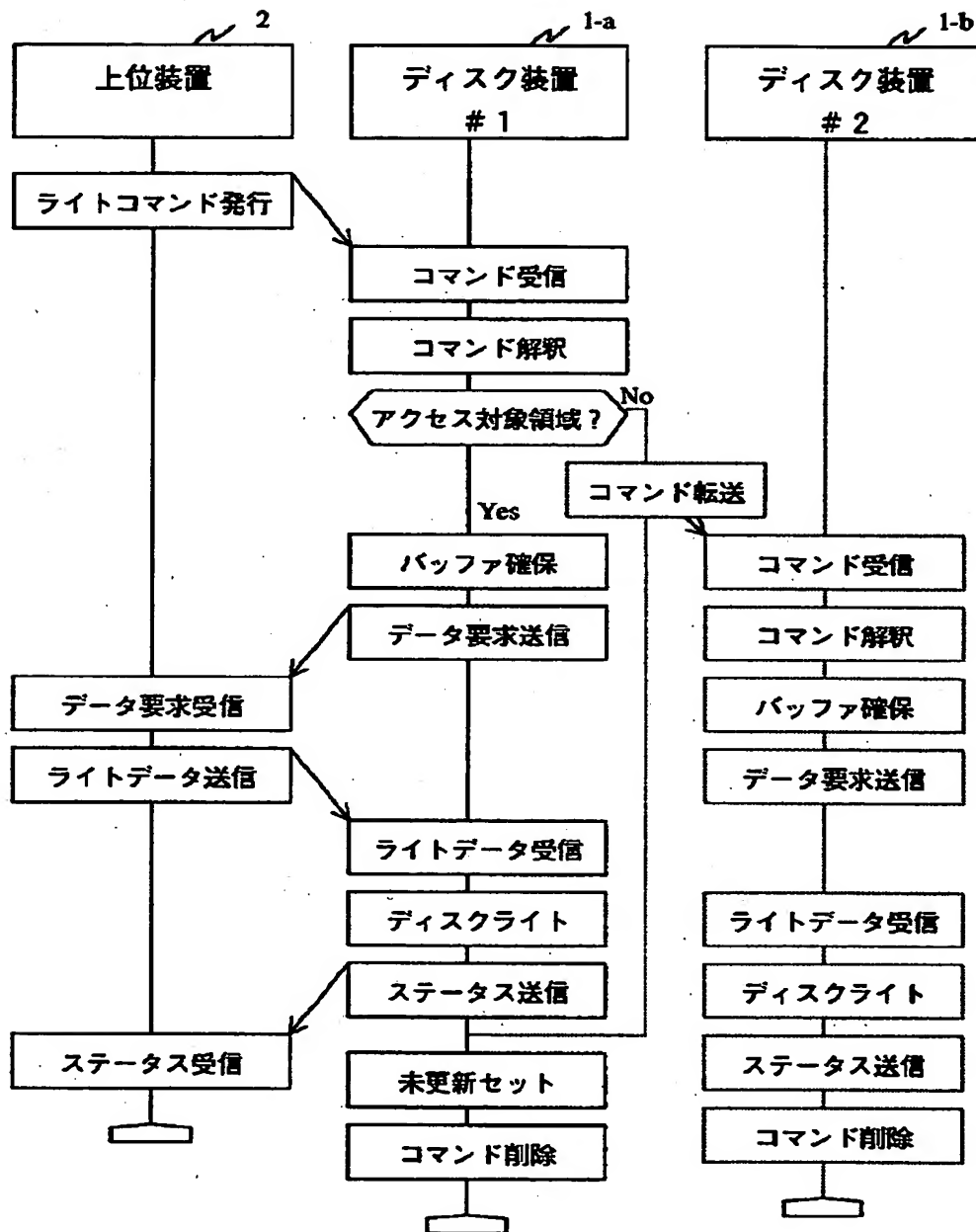


【図 11】



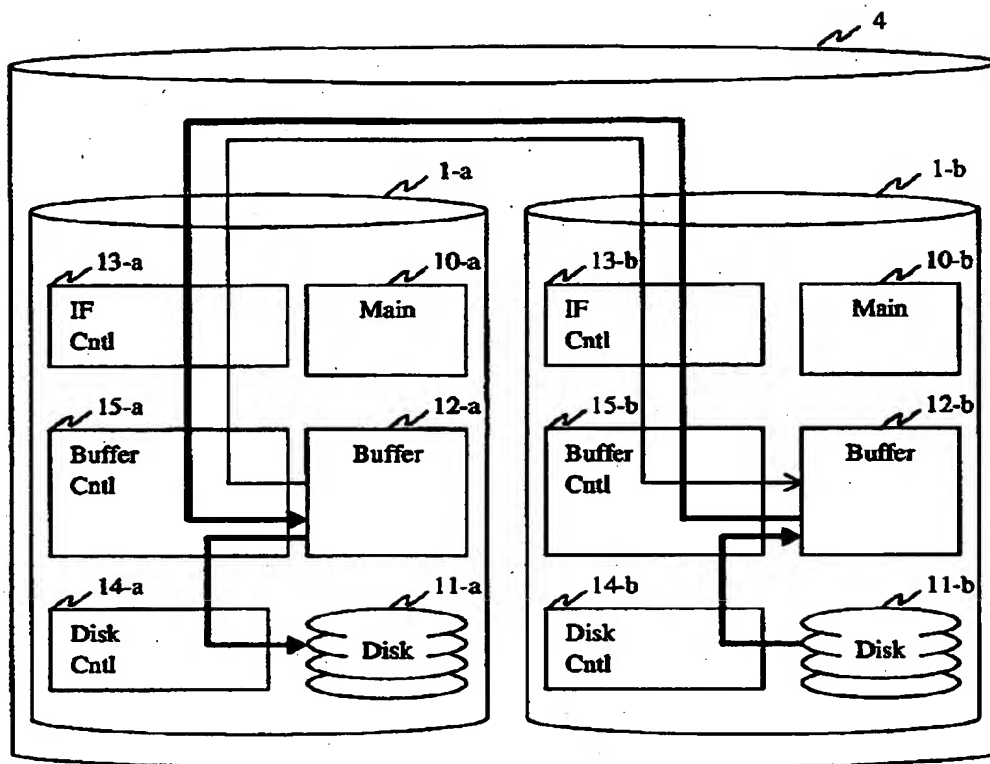
【図 1 2】

図 1 2



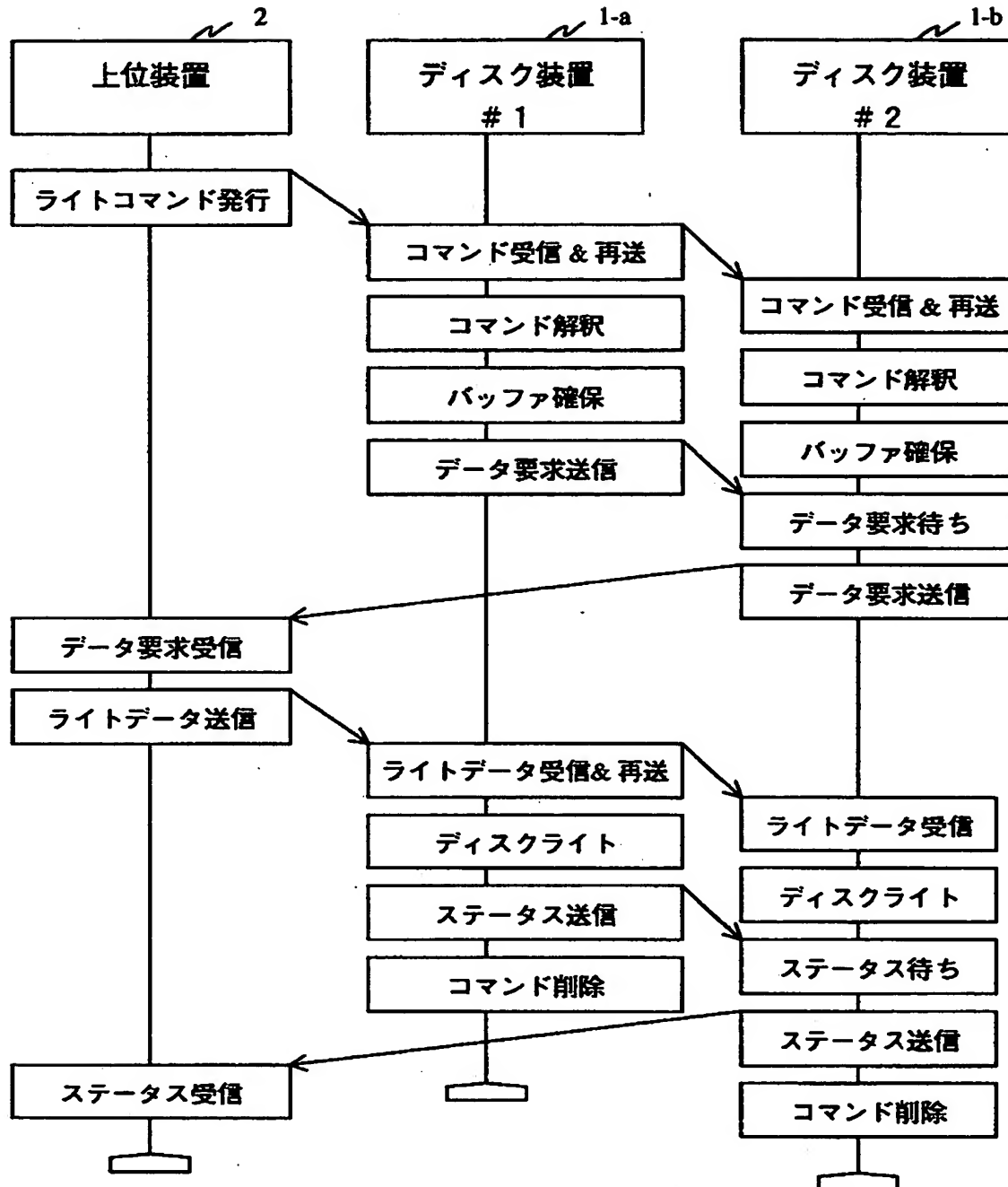
【図13】

図13

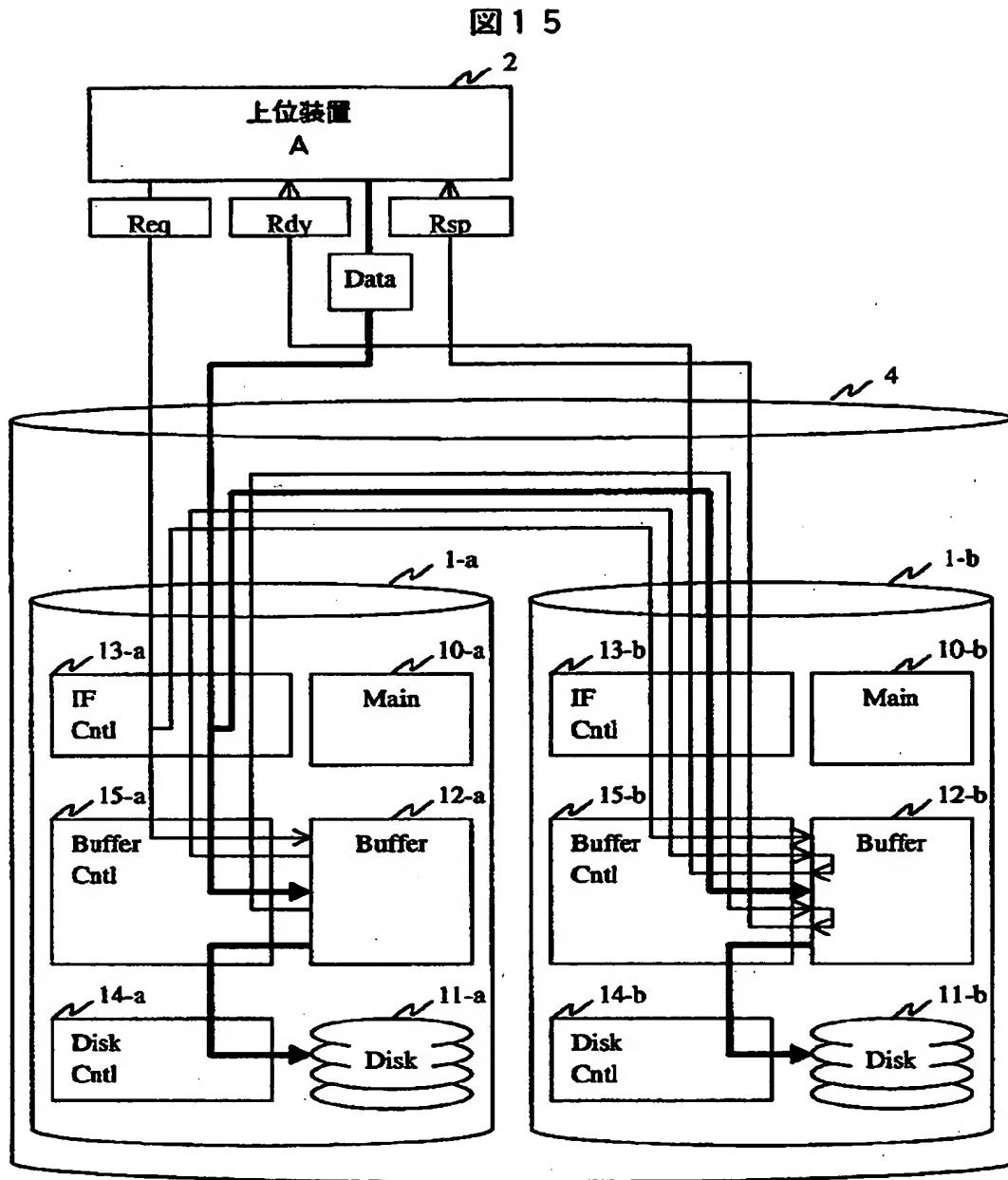


【図14】

図 1 4



【図15】



【図 16】

図 16

協調制御情報項目		記憶装置 # 1 設定例	記憶装置 # 2 設定例
記憶装置システムID (共有ID)		ID_0	ID_0
記憶装置ID (固有ID)		ID_1	ID_2
協調装置ID (協調記憶装置固有ID)		ID_2	ID_1
協調処理 (コマンド転送) モード		コマンド再送	コマンド受領
領域 管理 情報	協調処理領域	0 - (N-1)	0 - (N-1)
	アクティブ領域開始アドレス	0	N/2
	アクティブ領域サイズ	N/2	N/2

【図 17】

図 17

Bit Byte	7	6	5	4	3	2	1	0
0 - 7	FCP_LUN							
8	Command Reference Number							
9	Reserved					Task Attribute		
10	Task Management Flags							
11	Reserved						Rd Data	Wr Data
12	Operation Code							
13	Logical Unit Number			Flag				
14 - 17	Logical Block Address							
18	Reserved							
19 - 20	Transfer Length							
21	Control							
22 - 27	Reserved							
28 - 31	FCP_DL							

【図 18】

図 18

Byte Word	3	2	1	0
0	R_CTL	D_ID		
1	CS_CTL	S_ID		
2	TYPE	F_CTL		
3	SEQ_ID	SEQ_ID	SEQ_CNT	
4	OX_ID		RX_ID	
5	Parameter			

【図 19】

図 19

#	Host ID (S_ID)	OX_ID	RX_ID
0	ID_3	OX_00	RX_00
1	ID_3	OX_01	RX_02
2	ID_4	OX_01	RX_04
.	-	-	-
N	-	-	-

【書類名】 要約書

【要約】

【課題】

複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、上位装置からのアクセス要求を、該記憶装置システムを構成する複数の記憶装置間で協調して処理する制御方式を提供する。

【解決手段】

RAID1の記憶装置システムを構成する個々の記憶装置が、上位装置からのアクセス要求を共有する手段と、上記共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべき記憶装置に転送する手段とを具備する。

【効果】

簡単な構成で高性能なRAID1の記憶装置システムを実現する。

【選択図】 図 2

特2001-153345

認定・付加情報

特許出願の番号	特願2001-153345
受付番号	50100737476
書類名	特許願
担当官	第七担当上席 0096
作成日	平成13年 5月24日

<認定情報・付加情報>

【提出日】	平成13年 5月23日
-------	-------------

次頁無

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所